Self Organizing IP Multimedia Subsystem

Ashutosh Dutta, Christian Makaya, Subir Das, Dana Chee, and Joe Lin
Telcordia Technologies, NJ, USA

Satoshi Komorita, Tsunehiko Chiba, and Hidetoshi Yokota
KDDI Labs, Japan
Henning Schulzrinne
Columbia University, New York, NY

***Abstract*—** While there have been tremendous efforts to develop the architecture and protocols to support advanced Internet-based services over 3G and 4G networks, IMS is far from being deployed in a wide scale manner. Effort to create an operator controlled signaling infrastructure using IP-based protocols has resulted in a large number of functional components and interactions between those components. Thus, the carriers are trying to explore alternative ways to deploy IMS that will allow them to manage their network in a cost effective manner while offering the value-added services. One of such approaches is self organization of IMS. The self organizing IMS can enable the IMS functional components and corresponding nodes to adapt them dynamically based on the features like network load, number of users and available system resources. This paper introduces such a self organizing and adaptive IMS architecture, describes the advanced functions and demonstrates the initial results from the prototype test-bed. In particular, we show how all IMS functional components can be merged and split among different nodes as the network demand and environment change without disrupting the ongoing sessions or calls. Although it is too early to conclude the effectiveness of self organizing IMS, initial results are encouraging and it may provide additional incentives to the operators for network evolution.

Keywords; Self organizing, IMS, Adaptive, Context-aware

I. Introduction

The current version of IMS (IP Multimedia Subsystem) has several shortcomings that may act as deterrent for wider deployment. Depending upon the architecture there may be a need to have one-to-one mapping between functional component and a physical node where each node is equipped only to perform a certain function. This would require the deployment of a large number of network nodes which in turn will make the network management and operation difficult in particular the environment where dynamic adaptation is required based on the functionalities, processing, network loads and node failures. One way to circumvent these issues is to allocate redundant network resources but that will not help the service providers to achieve the goal of offering services at a lower cost. Alternatively, one can define new techniques that support dynamic adaptation with the nodes merging and splitting the IMS functional components. By node merging, we mean all the IMS functional components merge and operate on a single physical node; whereas, by node splitting we mean the functional components get distributed across different physical nodes. In its current form, IMS architecture and protocols do not have the mechanisms that can easily help IMS functionalities (e.g., P-CSCF, I-CSCF, S-CSCF) to easily migrate from one node to another node. Thus, it needs additional features and mechanisms to support self organizing capability. These additional features will enable the IMS network to adapt and distribute the functionalities based on the network condition and operational environment. Currently, IEEE NGSON (Next Generation Service Overlay Network) [1], ATIS SON [2] and ITU-T SG13 [3] are working on defining the requirements that would enable the IMS to support several features such as self organization, context awareness, and peer-to-peer services. These additional standards may help service providers to better manage their IMS network while reducing the cost of operations.

978-1-4244-4793-0/09/$25.00 © 2009 IEEE

This paper proposes and develops methodologies that are needed to support several self organizing features of IMS. These methodologies can provide additional flexibilities to the current IMS architecture and help service providers to better manage the IMS components in a dynamic environment where nodes or links are prone to failure and the system grows or sinks with the demand. This paper also describes the laboratory prototype test-bed and provides some initial results to demonstrate the feasibility of such a system.

The rest of the paper is organized as follows. We introduce related work and standards efforts in Section II. Section III describes the self organizing architecture and the associated functions that are needed to support this architecture. Section IV describes the operations of different nodes in the architecture. Section V shows call flows associated with different configuration scenarios, the test-bed prototype and initial results. Finally, Section VI concludes the paper.

II. Related work

The concept of self-organizing IMS networks is relatively a new topic and has not been widely studied. It is important to understand that the self-organizing IMS is different than P2P-SIP concept [4] where significant research results are available. Bessis [5] describes performance analysis and benefits of running multiple SIP servers on the same host. That paper shows how to design the IMS networks in order to maximize IMS server co-location and explains which types of SIP calls can benefit from the co-location of IMS servers. Fabini et al. [6] describe a minimal optimal IMS configuration with respect to architecture and QoS aspects. A virtual IMS test-bed (i.e., any IMS component is assigned its own virtual host), with different domains has been setup on one physical machine. It demonstrates the feasibility of an IMS system implementation within a single device (all-in-one). Matus et al. [7] propose a distributed IMS architecture by representing network functional elements in Distributed Hash Tables (DHT) overlay networks. The main focus was to distribute S-CSCF, I-CSCF and HSS functionalities by using an overlay network where these functionalities are merged in one node (called IMS DHT). However, none of these papers have looked into methods of supporting self organizing capabilities of IMS. Furthermore, these papers did not take into account reconfiguration of the functions when, for example nodes fail or for load balancing reasons.

Most recently, the Next Generation Service Overlay Network (NGSON) [1], a standards activity launched by the IEEE Communications Society in March 2008, is specifying a framework for service overlay networks. This framework will include context-aware, dynamically adaptive, and self-organizing networking capabilities including advanced routing and forwarding schemes that are independent of underlying transport networks. The overall goal is to make all IP service networks more autonomous and intelligent in the way that both service support and operation support can be improved significantly in these networks.

This paper works around the basic concept of NGSON that is based on self-organizing IMS. In particular, this paper develops the mechanisms and framework to facilitate self-organizing capabilities such as self-configuration, self-optimization and self-recovery at the service layer. It describes different functions that are needed to support the self organizing features of IMS as being pursued within IEEE NGSON.

III. SELF ORGANIZING IMS ARCHITECTURE

Figure 1 shows a specific deployment scenario for the self organizing IMS where nodes and links may fail. Figure 2 shows how an IMS node in Figure 1, capable of running several IMS functions, can adapt itself according to the server load and network conditions. It is assumed that all the nodes are capable of taking on the roles of the IMS components, such as P-CSCF (Proxy Call Session Control Function), S-CSCF (Serving Call Session Control Function), I-CSCF (Interrogating Call Session Control Function). These nodes can run one or multiple instances of different IMS functions. In this section, we introduce the functional behavior and interactions of different logical entities that constitute the self organizing IMS architecture.

Self-Organizing Approaches

Self-organizing IMS can be based on one of the two modes: *centralized* or *distributed*. In the centralized mode, there is a master node that maintains a database with operator policy and state information for all nodes under its control. For example, this master node can be the HSS (Home Subscriber Server) as specified in [8]. The reason behind this choice could be that HSS is the master database of the cellular network and the availability and reliability of such database is much higher than ordinary nodes.

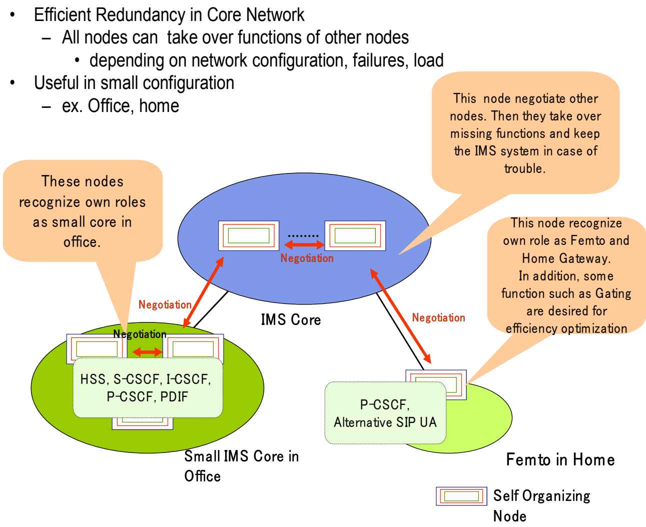

Figure 1: Application of self organizing IMS

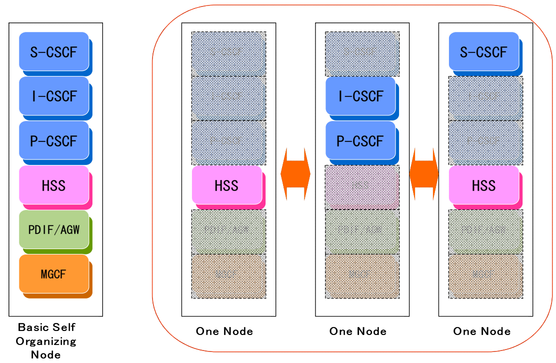

Figure 2: Basic concept of self organizing IMS

The Master Node database is updated when i) a new node notifies the Master Node about its capabilities; ii) a specific IMS role has been assigned to the node, and, iii) a node changes roles due to overloading or failure. An efficient policy-based mechanism should be defined in the master node to assign the functionalities or roles.

In the distributed mode of operation the new node announces its presence through a multicast message. Existing IMS nodes reply to this message if they need to transfer some of their current functionalities. This may result in conflicts but it can be resolved by a simple rule such as, first in first out (FIFO) based on response message. If a node receives no responses, it should assume that there exists no other node and it takes over all roles.

In either the centralized or distributed mode, negotiation or capabilities exchange and event notification protocols should be defined. In fact, the decision to assign a role to a new node will be based on its capabilities (e.g., CPU, load, processing power, memory). We focus on the centralized mode in this paper. Some of these functionalities can be easily applied to a distributed mode as well. We next describe the important functions of self-organizing IMS that are needed in addition to the standard IMS functions.

Nodes discovery function

There are two kinds of node discovery function in Self Organizing IMS: 1) a master node discovers the IMS nodes and their capability 2) mobile node discovers the IMS node for registration. The master node discovers the other IMS nodes and their capability via a negotiation protocol. Mobile node on the other hand can use standardized mechanisms such as DHCP options [9] or DNS query [10] for IMS node discovery.

Role request and assignment

When an IMS node comes online, it requests a role or function from the master node by sending its capability. The capability may include IMS functionality such as P-CSCF, S-CSCF or I-CSCF, and other server specific capability such as processing power, memory and so on. Based on this capability, the master node assigns a given role to the new node. The role assignment can happen due to load balancing, node failure and other node management purposes that can be governed by high level policies configured at the master node. Alternatively, it can be network event-based such as link failures or sudden change in bandwidth. When such event occurs, master node notifies the appropriate nodes and the network will be auto-configured accordingly. The network auto-configuration can happen in several ways. For example, an IMS node can notify the mobile node about a role change provided the mobile node subscribes to an event after registration. This event could be defined as a function that indicates the change of role for an IMS node. SIP Event notification messages such as SUBSCRIBE and NOTIFY [11] may be used for this event change notification. The role assignment algorithm on the other hand can be implementation specific.

Protocol interaction between nodes

To allow deployment of self-organizing IMS networks, protocol interactions among different IMS nodes are necessary. These interactions are among IMS nodes and mobile nodes, and between the master node and IMS nodes.

Node monitoring

To allow the Master Node to determine the functional behavior of all other nodes in the network, a periodic message (heartbeat) will be sent to all known nodes. If a response is not forthcoming within a specific time, the given node is marked as troubled or failed, and the master node will distribute its functionality to the remaining nodes dynamically.

IV. SELF-ORGANIZING IMS NODE FUNCTIONS

In this Section, we describe the functionality supported by each node involved in self-organizing IMS network.

Operations of the Master Node

The master node is the main component of the centralized approach. In addition to assigning roles to nodes, it has other functions depending on the type of event. As an example, in Figure 3, if the P-CSCF (e.g., P1) role changes, the Master Node notifies all S-CSCF nodes and provides them the information about the new P-CSCF. When the S-CSCF receives this notification, it establishes a list of mobile nodes assigned to this P-CSCF and notifies them by way of the new P-CSCF (e.g., P2) (Receive Role Change). Mobile node (MN) and User Equipment (UE) are used interchangeably and have the same meaning in this paper. The notification message sent by S-CSCF to the mobile nodes contains information about the new P-CSCF. Upon receipt of notification from the S-CSCF, mobile nodes re-register to the new P-CSCF and re-subscribe to event state change for future changes.

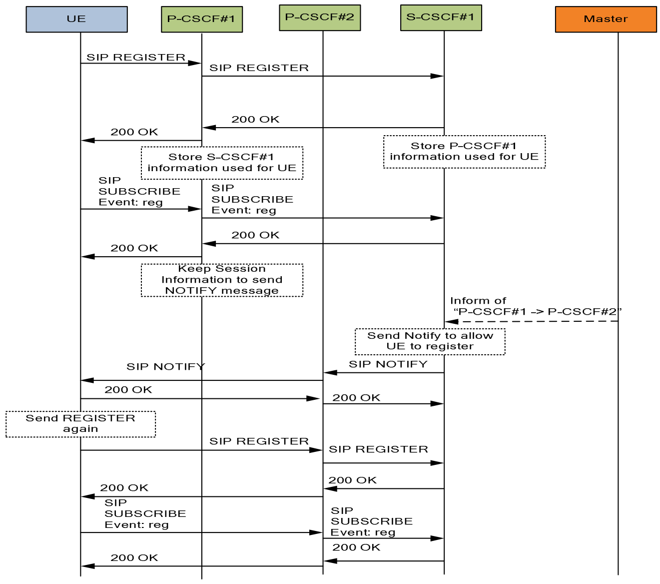

Figure 3: Role of master node during P-CSCF change

On the other hand, if the S-CSCF role changes, the master node notifies all P-CSCF followed by notification of mobile nodes. The notification message includes information about the new S-CSCF. Upon receipt of this notification message, the mobile node re-registers and subscribes to event state change. To handle load balancing, the master node may request each of the IMS components to support a given percentage of mobile nodes previously attached to the IMS node with role change functionality.

Operations of mobile node

The operation of mobile node is based on SUBSCRIBE/NOTIFY methods. The mobile node should be able to receive and process the notification message (NOTIFY) received either from S-CSCF when P-CSCF role changes or from P-CSCF if S-CSCF role changes. Upon receipt of this NOTIFY message, the mobile node uses such information to register and subscribe again to event change. Mobile node's ongoing session including the QoS of the media is not affected since the access network that the mobile node is attached to is not changing. However, if the mobile node needs to establish a new session, it should include the Replace-header field in the INVITE message or use re-INVITE message.

Operation of the P-CSCF

To allow self-organizing IMS networks, P-CSCF components should be able to support the following additional operations:

- Receive notification of role change from the master node;
- Proxy and store mobile node's subscription information during SUBSCRIBE operation;
- Send NOTIFY message with information about the new S-CSCF;
- Retrieve all mobile nodes registered to S-CSCF upon receipt of event change from the Master Node;
- Process mobile node's re-registration request after a role change event.

Operation of the S-CSCF

To allow self-organizing IMS networks, S-CSCF components should be able to support the following additional operations:

- Receive notification of role change from the master node;
- Store mobile node's subscription and profile information;
- Send NOTIFY message with information about the new P-CSCF;
- Retrieve all mobile nodes registered to the P-CSCF upon receipt of event change from the Master Node;
- Update HSS with the mobile node information after the role change.

V. IMPLEMENTATION AND TESTBED PROTOTYPE

In this section, we first illustrate a few migration scenarios of functional components to achieve a self organizing IMS network and then describe the associated call flows and proof-of-concept from a prototype test-bed. Ideally, self organizing IMS should be able to support all kinds of reconfiguration in the network as shown in Figure 4. For example, Scenario 1 in Figure 4 shows that all the IMS components are functioning in one node, N1. Scenario 2 shows that P-CSCF functionality is migrated to node N2. Scenario 3 corresponds to a case when the P-CSCF functionality gets split into two nodes, P1 and P2. Scenario 4 depicts when the S-CSCF functionality is relocated to a different node. Scenario 5 corresponds to a case when the S-CSCF functionality gets split into two nodes S1 and S2, Scenario 6 shows I-CSCF's functionality getting migrated to a new node, and finally, the transition step (6) shows the merging functionality where all the functions merge with a single node. It is important to note that our goal is not to affect the end users (whether they have an ongoing session or initiate a new session) due to these changes in the configuration within the core of the network. In Figure 4, H stands for HSS, I-CSCF is termed as I, S-CSCF and P-CSCF are denoted as S-CSCF and P-CSCF, respectively.

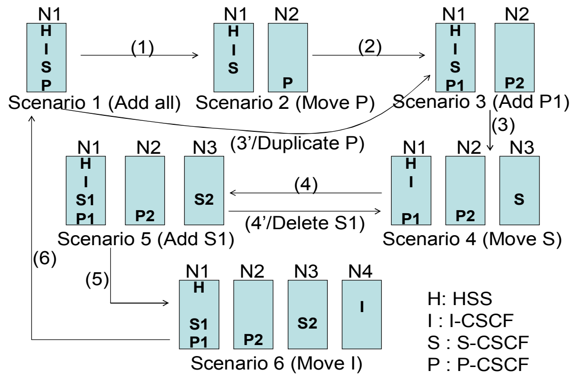

Figure 4: Possible re-configuration scenarios for IMS nodes

Scenario 1: HSS, S-CSCF, I-CSCF and P-CSCF are deployed in the same node

Figure 5 shows the call flows for scenario 1, where all SIP-based IMS components (i.e., HSS, S-CSCF, I-CSCF, and P-CSCF) are deployed in the same node. In this scenario, mobile node 1 (MN1) and mobile node 2 (MN2) register with Node 1 (N1) and a call session is established between them. To allow deployment of self-organizing IMS networks, after registration, mobile nodes must subscribe to S-CSCF and P-CSCF for the role change event.

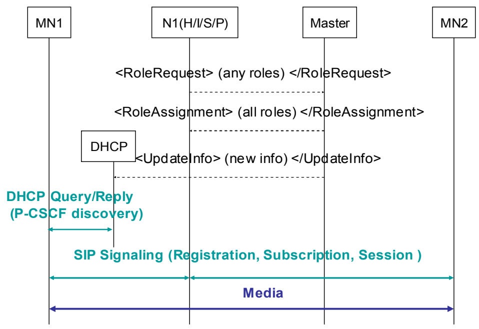

Figure 5: Call flow for scenario 1

We use MESSAGE method for any communication between the mobile node and the master node. Initial capabilities exchange between the IMS node and master node and role assignment functions are carried out using MESSAGE method.

Figure 1.

Figure 6 shows a sample MESSAGE method [12] showing "Role Request" and "Role Assignment". As part of role request, the IMS node offers its capabilities in terms of

CPU, memory, and load. Role Query response is not shown. Figure 7 shows the call flows for scenario 2 as depicted in Figure 4, which corresponds to the case when the P-CSCF functionality is relocated in a new node (i.e., Node 2 (N2)) while HSS, S-CSCF, and I-CSCF remain in Node 1. In this scenario, when Node 2 comes online, it exchanges its capability information with the Master Node and it is found that Node 2 has the ability of serving as P-CSCF.

```
Request-Line: MESSAGE
sip:node01@[3ffe:5::201]:5066 SIP/2.0
<RoleRequest>
  <Id
     address="3ffe:5::201"
     name="node01"/>
     <Ability
          cpus="1"
          load="0.00"
          memory="65.43"
          rate="2800.000"
          totalMemory="1000"/>
          <CurrentRoles/>
</RoleRequest>

Request-Line: MESSAGE sip:domain1-ims
rtr@[3ffe:5::100]:5066 SIP/2.0
eXtensible Markup Language
 <RoleAssignment>
   <Id
     name="node01"/>
    <database
          name="None"/>
    <CurrentRoles>
       <Role
          name="hss"/>
       </CurrentRoles>
 </RoleAssignment>
```

Figure 6: Role request and assignment using MESSAGE

Hence, the master node decides to transfer or move the P-CSCF functionality to Node 2 which acts now as P-CSCF for both mobile node 1 and mobile node 2. The procedure is described as follows:

- The old IMS Node (N1) is providing the IMS functionalities for the Mobile Node.
- When a new IMS Node (N2) comes online, it sends a request to the Master Node for role assignment. According to the capabilities provided in the role request response message, the master node decides the role assignment to N2 governed by the operator policy. The master node provides information about the current IMS entity if the role has been already assigned to other IMS nodes.
- Upon deciding on the role assignment, the Master Node notifies the S-CSCF (N1/S) about role assignment changes and provides information about the new IMS Node (N2/P). The S-CSCF will notify all mobile nodes registered to the previous P-CSCF.
- The new IMS Node (N2/P) retrieves information (state or context) from the old IMS Node (N1).
- In order to allow new mobile nodes to discover the correct P-CSCF, the Master Node updates the DHCP server configuration.

Scenario 3 corresponds to the situation when P-CSCF functionality gets split among two nodes (P-CSCF1 in Node 1 and P-CSCF2 in Node 2). In this scenario, the Master Node decides to assign P-CSCF role to Node 1 while earlier P-CSCF role is located in Node 2 (from Scenario 2). Mobile node 1 is assigned to Node 1 as its P-CSCF (i.e., P-CSCF1) while mobile node 2 is assigned to Node 2 as its P-CSCF (i.e., P-CSCF2). After these steps, a call session is established between mobile node 1 and mobile node 2. Note that the P-CSCF functionality splits happen after the initial registration of mobile node 1 and mobile node 2. In other words, mobile node 1 and mobile node 2 were associated to the same P-CSCF (Node 2).

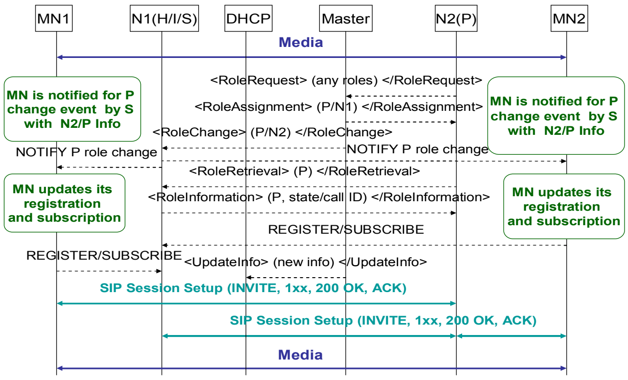

Figure 7: Call flow for P-CSCF relocation (Scenario 2)

Figure 8 shows the call flows for scenario 4 that corresponds to the situation when S-CSCF functionality is relocated in a new node (i.e., Node 3). In this scenario, when Node 3 comes online, it exchanges its capability information with the Master Node and it is found that Node 3 is suitable for serving as S-CSCF. Then, S-CSCF functionality of Node 1 is transferred to Node 3 and Node 1 (i.e., P-CSCF, HSS or I-CSCF) is notified for this change.

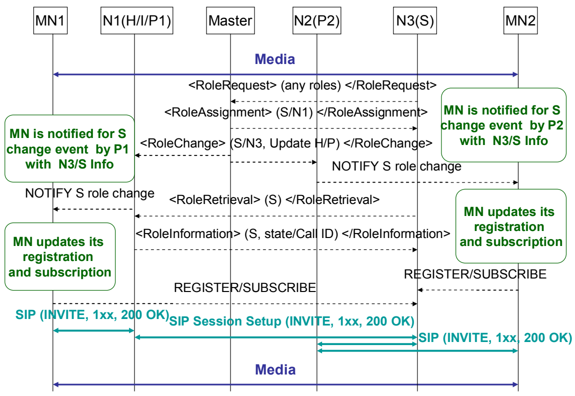

Figure 8: Call flow for S-CSCF relocation (Scenario 4)

Transition step (6) depicts a situation when IMS main nodes (P-CSCF, S-CSCF) functionalities move back to one node due to failure of other nodes or load balancing reasons. The Master Node detects this state change or event and restores all functionalities previously supported by each IMS node to Node 1. Mobile node 1 and mobile node 2 re-register with Node 1 again. Before making such merging decision in one node, Master Node needs to consider several factors such as, time to restore the functionalities, service disruption.

Figure 9 shows the functional components in our experimental test-bed. All components in the test-bed are Linux-based and there are two subnets (access networks) and one home network. The home network is equipped with all SIP-based IMS components: HSS, I-CSCF, S-CSCF and P-CSCF. All IMS functionalities can run in the same node. The master node is located in the router that also acts as the DNS server as well as DHCP server. The edge routers act as 3GPP's PDIF (Policy Decision and Information Function) and DHCP relay agent. We use "*dibbler*" as a DHCP client on the mobile. P-CSCF discovery procedure is based on DHCP as specified by standard IMS. We have used XML-based query-response mechanism to obtain the required information from the master node and for role assignment.

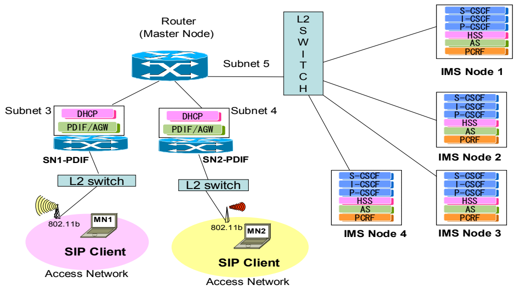

Figure 9: Overview of self-organizing IMS test-bed

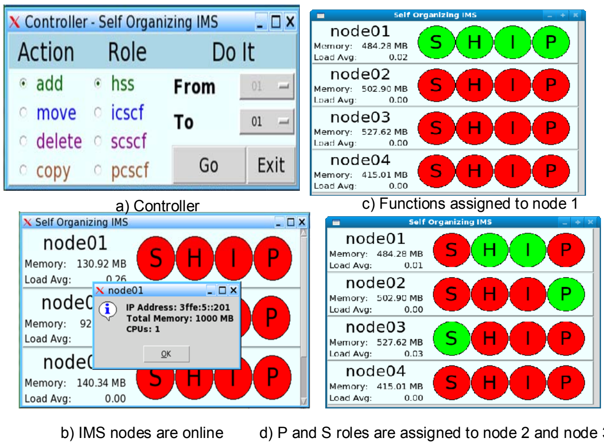

Figure 10: Screen shots from self organizing IMS

SIP stack used in the test-bed is based on NIST implementation [13]. Mobile nodes use SIP user agent based on SIP Communicator [14] to communicate with the IMS nodes. Mobile nodes communicate with the DHCP server that resides in the router via the DHCP relay agent in the edge router in order to obtain the IP address and discover SIP servers. Figure 10 shows the screen shots of different operations in the test-bed. As shown in Figure 10, the green illustrates that the specific functionality (e.g., P-CSCF, S-CSCF, I-CSCF) in the current node is active. Red indicates that the nodes are inactive. It also shows how initially all the IMS functions are running in Node 1 but some changes in network condition migrates some of these functionalities to Node 2 (e.g., P-CSCF) and Node 3 (e.g., S-CSCF). Preliminary performance results for server reconfiguration indicate that it takes less than one second to re-assign the respective server functionalities by the Master Node. This reconfiguration time plays an important role in determining the extent of service interruption. We are implementing both UE (mobile)-involved and UE (mobile)-non-involved cases to realize the self organizing IMS. UE-involved case requires that mobile is involved in signaling exchange during reconfiguration process. UE-non-involved case does not require mobile's involvement during network node reconfiguration, network entities such as load balancer and SIP proxies help re-registering the mobile when the network components change their roles.

V. CONCLUSIONS

We believe that self organizing IMS is an important step towards the successful deployment of IMS networks. A policy-based configuration and reconfiguration of IMS components will help operators reduce the cost and complexities of current networks. While there are standards groups such as IEEE NGSON, ATIS SON and ITU-SG13 are currently working on the requirements and architecture of self organizing, context-aware networks, this paper goes one level beyond by realizing the concept of self organizing IMS in a laboratory environment. Our approach is self adaptive in the sense that we preserve the basic IMS node functionalities with network and operational environment change. Our future research includes implementing two scenarios: i) where mobile node is aware of reconfiguration and participates during self organizing process, ii) where mobile node is not aware of self organizing IMS and does not participate during the reconfiguration process.

REFERENCES

[1] IEEE NGSON: http://grouper.ieee.org/groups/ngson

[2] ATIS SON: http://www.atis.org/SON

[3] ITU-T SG13: http://www.itu.int/ITU-T/studygroups/com13/index.asp

[4] E. Marocco et al., " Interwroking between P2PSIP Overlays and IMS Networks – Scenarios and Technical Solutions," www.p2psip.org

[5] T. Bessis: Improving Performance and Reliability of IMS Network by Co-Locating IMS Servers, Bell Labs Technical Journal, Vol. 10, No. 4, pp. 167-178, 2006.

[6] J. Fabini et al., "IMS in a Bottle," Initial Experiences from an OpenSER-based Prototype Implementation of the 3GPP IP Multimedia Subsystem, in Proc. of the Int'l Conference on Mobile Business (ICMB'06), June, 2006, Copenhagen.

[7] M. Matuszewski, M. Garcia-Martin: A Distributed IP Multimedia Subsystem (IMS), IEEE Intl. Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM'07), pp. 1-8, 2007.

[8] 3GPP: Third Generation Partnership Project (www.3gpp.org).

[9] H. Schulzrinne, "DHCP Option for SIP Servers," RFC 3361, IETF, August 2002.

[10] A. Gulbrandsen et al., "DNS RR for specifying the location of services," RFC 2782, IETF, Februry 2000.

[11] A.B Roach, SIP Specific Event Notification, RFC 3265, IETF, June 2002.

[12] B. Campbell et al., "Session Initiation Protocol (SIP) extension for Instant Messaging," RFC 3428, IETF, December 2002.

[13] NIST implementation (URL): https://jain.sip.dev.java.

[14] SIP Communicator (URL): http://sip-communicator.org.